CR-143157

E7.5-10.3.5.5.

ENVIRONMENTAL RESEARCH INSTITUTE OF MICHIGAN

FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN

P. O. BOX 618 • ANN ARBOR • MICHIGAN • 48107

PHONE (313) ~~XXXXXXX~~ 994-1200

Information Systems and Analysis

101900-59-L
15 July 1975

Developing Processing Techniques for Skylab Data
Monthly Progress Report, June 1975

(E75-10355) DEVELOPING PROCESSING TECHNIQUES FOR SKYLAB DATA Monthly Progress Report, Jun. 1975 (Environmental Research Inst. of Michigan) 11 p HC $3.25 CSCL 05B

N75-28493

Unclas
G3/43 00355

EREP Investigation 456 M
NASA Contract NAS9-13280

"Made available under NASA sponsorship in the interest of early and wide dissemination of Earth Resources Survey Program information and without liability for any use made thereof."

Prepared by

Richard F. Nalepka - Principal Investigator
William A. Malila - Co-Principal Investigator
James P. Morgenstern

NASA Technical Monitor

Mr. Larry B. York/TF6
National Aeronautics and Space Administration
Johnson Space Center
Principal Investigator Management Office
Houston, Texas 77058

Developing Processing Techniques for Skylab Data
Monthly Progress Report, June 1975

The following report serves as the twenty-eighth monthly progress report for EREP Investigation 456 M which is entitled "Developing Processing Techniques for Skylab Data". The financial report for this contract (NAS9-13280) is being submitted under separate cover.

The purpose of this investigation is to test information extraction techniques for SKYLAB S-192 data and compare with results obtained in applying these techniques to LANDSAT and aircraft scanner data.

During the reporting period progress was made in the continuing analysis of the effects of spatial misregistration of SDOs on boundary pixels and in the implementation of the mixtures classifier. Additionally processing was begun on the signature extension task.

Effects of Channel-to-Channel Spatial Misregistration on Classification Accuracy and on Proportion Estimation

Previous reports have described simulation techniques developed to investigate effects of channel-to-channel misregistration and experiment designs to aid in the analysis of any effects. A discussion has been presented on the effects of spatial misregistration on field-center classification accuracy. During this report period further data processing stages of the experiment were carried out, and an analysis was made in an effort to study the effects of channel-to-channel misregistration on border (mixture) pixels. A discussion of this analysis follows.

The study conducted centered on the analysis of four basic problems: (1) the effect of misregistration on the classification of a mixture pixel of two ground covers; (2) the effect of misregistration on the false alarm rate of any given crop among mixtures of two other ground covers; (3) the effect of misregistration on proportion estimation; and (4) the effect of misregistration on the proportion of field center pixels in the data. In this report we shall deal with the first three analyses. The fourth problem will be discussed next month.

In order to study these four problems, the experiment described in the April monthly was carried out. Two types of signature simulations were required. First, signatures representing field center distributions misregistered for factors of 1/3, 2/3 and 1 whole pixel in SDOs 2, 12 and 17 were calculated. New distributions representing mixtures of all permutations of two ground covers for varying proportions were simulated as follows. Let α_{iA} and α_{iB} be the proportions of distributions A and B in the i^{th} channel used to simulate a mixture of ground covers A and B. For perfectly registered signatures, α_{Ai} was set to 2/3, 1/3 and 0 for every channel i. However for misregistered signatures, the channels out of registration would be in different proportions. For example, if a signature was misregistered by 1/2 a pixel the proportion of cover type A would be $\alpha_A - 1/2$. Hence any field-center pixels in the registered case within 1/2 pixel of the boundary would become mixture pixels in the misregistered case. (In effect there would be fewer field center pixels.) Therefore signatures representing mixtures of misregistered distributions were simulated with proportions of α_{iA} and α_{iB} in the registered channels i and $\alpha_{jA} - \beta$, $\alpha_{jB} + \beta$ in the misregistered channels j where β is the degree of misregistration.

Once the simulated signatures were attained the program PEC was run to calculate the expected performance for each set of signatures representing a given misregistration. That is, given the best linear decision boundaries between the 5 field-center signatures, what will the expected classification of mixture pixels appear to be. It should be mentioned at this point that processing was also carried out with only one channel misregistered, but analysis has not been completed and only the analysis of the three-channel case will be presented.

In order to simplify the discussion of the results, the presentation here will center only on the effects noted for brush and grass mixture pixels interacting with brush, grass and corn signatures, where the channels were assumed misregistered from the brush into the grass ground cover. This example was chosen for presentation here because corn, grass and trees comprise almost three-fourths of the scene. It represents neither one extreme nor the other — it is in fact fairly typical of the whole study. What will be particularly noted is (1) the effects on the classification of brush-grass mixtures as either brush or grass, (2) the corn false alarm rate, and (3) the effect on proportion estimation. Three sets of curves will be presented. These curves display the expected performance of brush and brush-grass pixels as a function of the proportion of each crop type. In a sense one could envision, as an aid in studying these graphs, a resolution element moving across a fixed field boundary and at various locations the expected probability of that resolution element's classification would be calculated. Note in each of the following graphs a zone representing pure field center pixels in the registered case has been labelled as well as an area representing mixtures of varying degrees. The width of these zones is exactly one pixel and the field boundary would appear at the point (1/2, 1/2) on the ordinate.

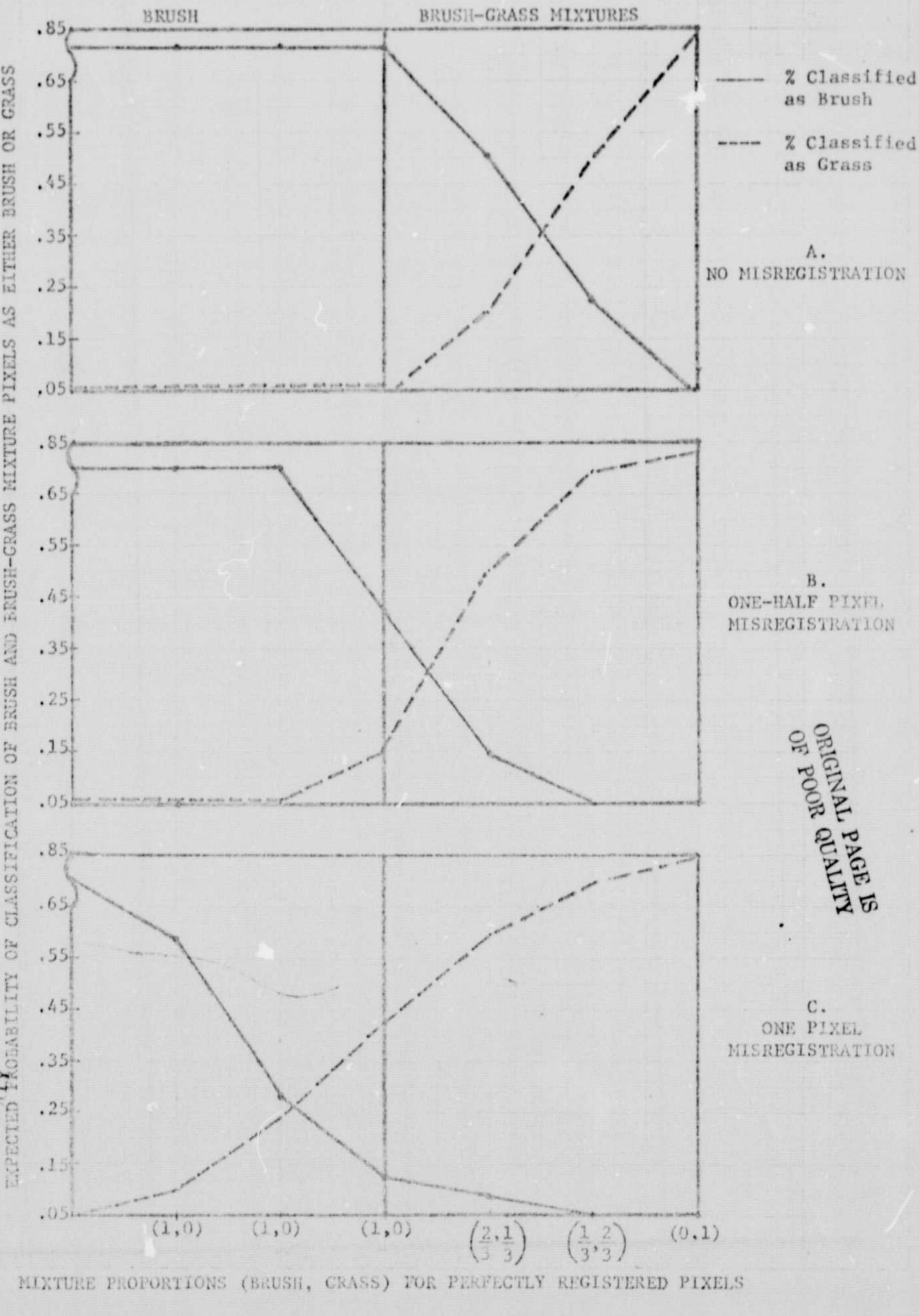
Figure 1 displays three graphs, one for each degree of misregistration considered, plotting the expected probability of classifying brush and brush-grass mixtures as brush (the solid line) or grass (the dashed line). In Figure 1-A, on top, one notes that in the area designated brush, these field center pixels are for the most part classified as brush. As the mixture of brush and grass becomes predominantly grass, the performance curve increases for grass and decreases for brush. Comparing Figure 1-A with Figures 1-B and 1-C (i.e., as misregistration increases), one notices immediately that the field center brush pixels are not recognized as brush with as much consistency. The expected performance of those most near the border deteriorates from around 80% to about 15%. This indicates that misregistration does effect the correct classification of field center pixels to a significant degree.

Figure 2 displays the expected probability of classifying a brush or brush-grass pixel as corn. Even in the registered case the corn false alarms among brush-grass pixels are significant. As misregistration is introduced, more and more corn false alarms occur among pixels that were pure field center brush pixels in the registered case. In fact those most near the border are called corn with up to 40% regularity. In view of this graph alone one cannot dismiss the significant increase in corn false alarms introduced by misregistration of the data.

A question of obvious concern is to what extent proportion estimation is affected by these effects of channel-to-channel misregistration. It is argued generally that errors tend to compensate for other errors; that is, errors are made uniformly in all directions and over a large sample their affects will be cancelled. The surprising corn false alarm rate among registered pixels of brush-grass already makes the process of proportion estimation a less than exact science. The increased number of false alarms to be expected with the introduction of misregistration places even more reliance on compensating errors.

Figure 3 is presented to show that the errors introduced are not strictly compensatory for proportion estimation, especially when misregistration is introduced in the scene. Let us focus our attention on the estimation of the proportion of corn. Noting an increased rate of corn false alarms among brush-grass pixels, these would necessarily have to be compensated for by a decrease in the correct classification of corn or mixtures of corn-other pixels (here we use the expression correct classification in the sense that mixtures of two covers A and B are classified as either A or B). Figure 3 is a graph of the expected probability of "correct"

FIGURE 1. EXPECTED CLASSIFICATION PERFORMANCE OF BRUSH, BRUSH-GRASS MIXTURE PIXELS



ORIGINAL PAGE IS OF POOR QUALITY

FIGURE 2. CORN FALSE ALARMS AMONG BRUSH AND BRUSH-GRASS MIXTURE PIXELS

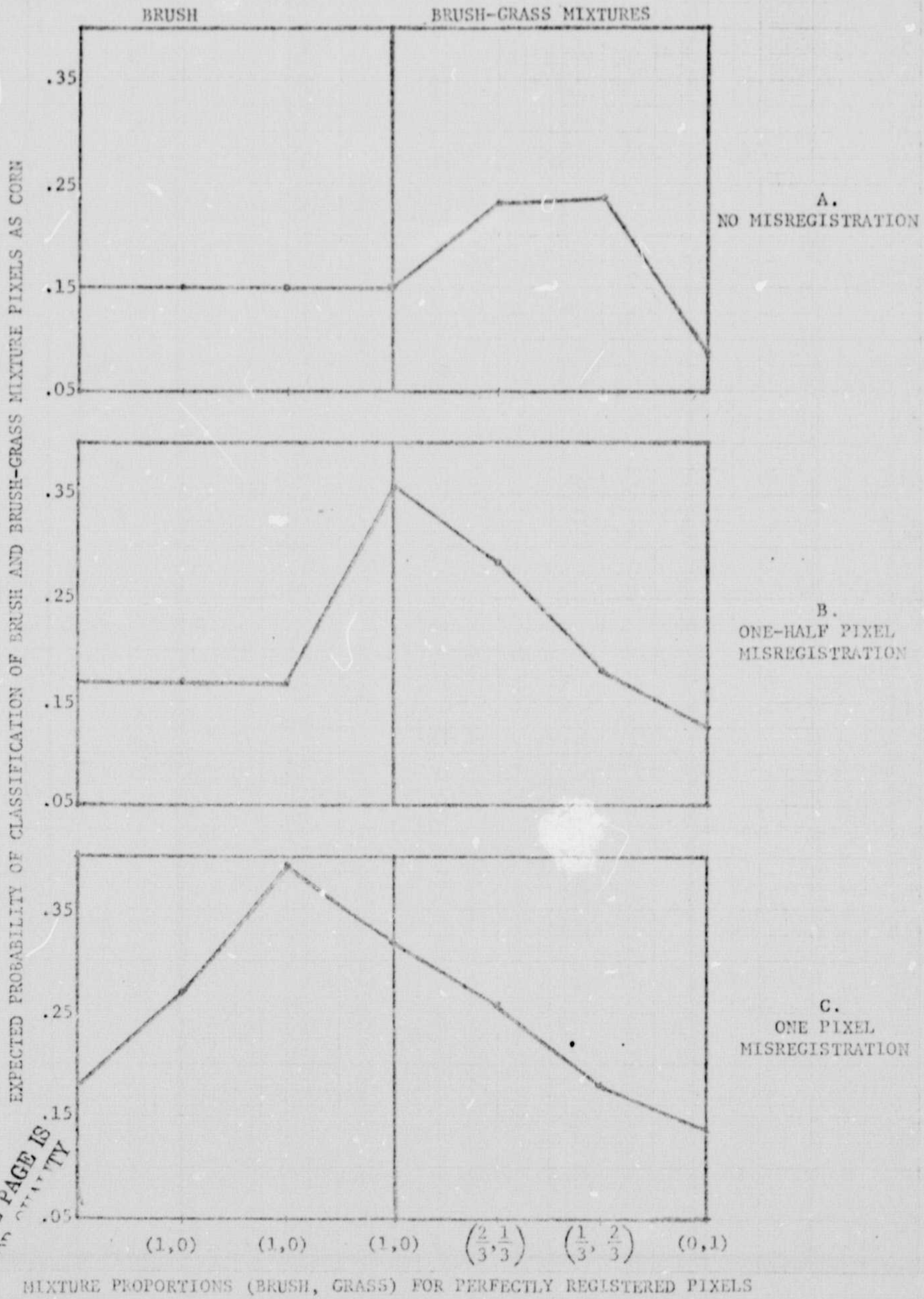
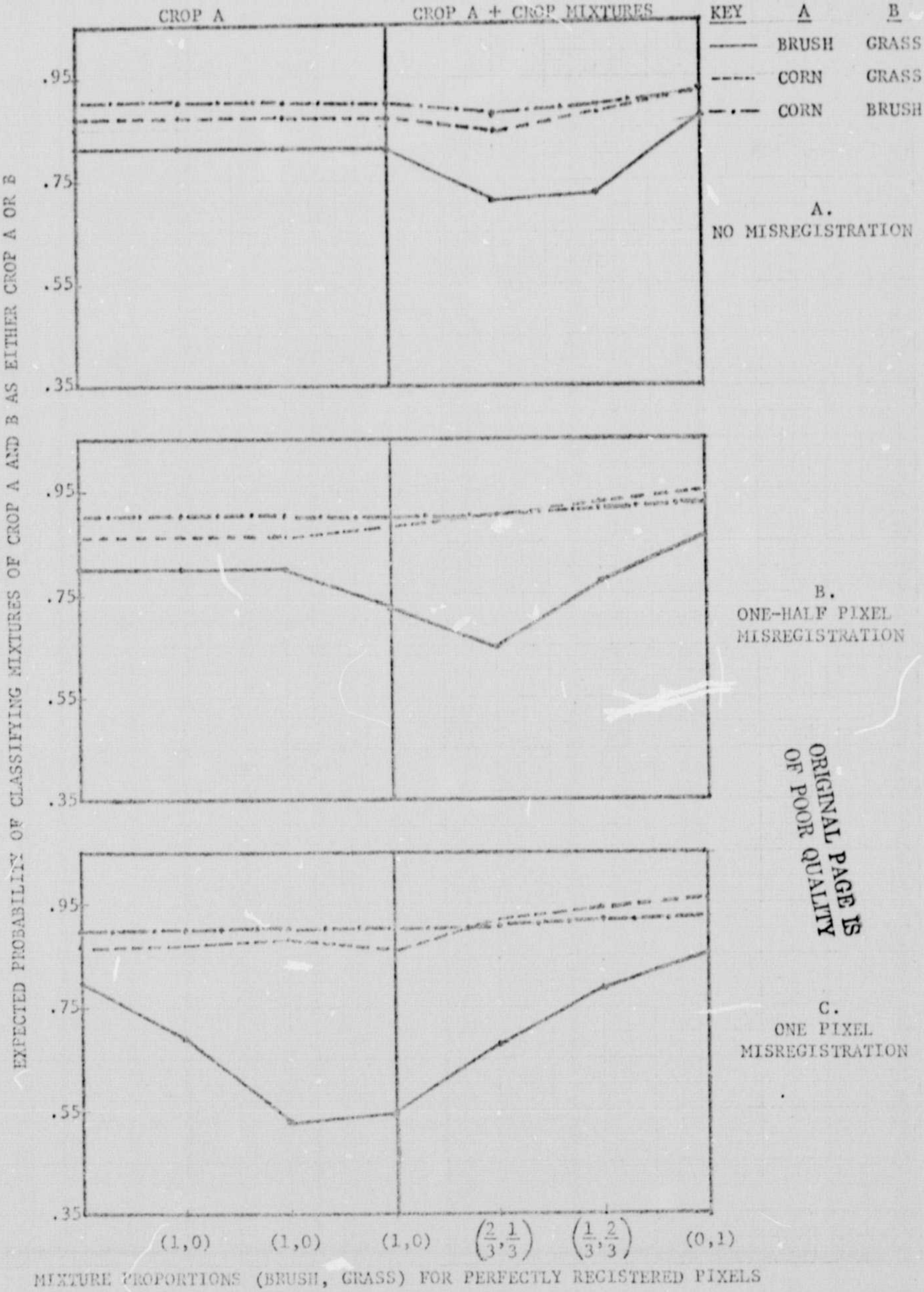


FIGURE 3. EXPECTED CLASSIFICATION OF MIXTURE PIXELS OF CROPS 'A' AND 'B' AS EITHER CROP A OR CROP B



ORIGINAL PAGE IS OF POOR QUALITY

ORIGINAL PAGE IS OF POOR QUALITY

classification of two ground covers as labelled as a function of the mixture proportion. The solid line indicates the amount of brush-grass correctly classified. With more misregistration there are more false alarms particularly of corn, as previously noted. However the correct classification of corn, corn-grass or corn-brush pixels does not correspondingly decrease indicating that corn may be overestimated in the given scene.

One can conclude from the evidence presented that the effects of misregistration are indeed significant in a detrimental manner in (1) the correct classification of border and near-border pixels; (2) in increasing the rate of false alarms, and (3) in proportion estimation.

Subresolution Element Processing

Last month we reported on initial geometrical analysis of the set of spectral signatures extracted from the data for use with the mixtures classifier. During the current period we continued by testing the chosen subset of signatures on a small portion of the data. Time was spent adapting the necessary software for use with S-192 data.

Before proceeding further, a short explanation of the manner in which the mixtures processor is applied is in order.[1]

It is obvious that a pixel may be purely or almost purely of one ground class, or it may be a mixture of several ground classes. Thus the algorithm used, as its first stage, determines the several likeliest possibilities. First, the most probable single signature for a pixel, and the attendant chi-square value are determined. Then the program determines the most probable pair of signatures, the proportion of each cover in the pixel, and an associated chi-square value. The pixel may be further analyzed as a mixture of three and four covers. For reasons of processing time and computer space requirements this study will limit the hypothesis to pure or two class mixture pixels. This is not an unrealistic restriction when one considers the scan swath over the ground. For an agricultural area like the current data set, most mixture pixels will occur at field boundaries such that the vast majority of mixture pixels will be mixtures of two ground classes. (In general, it is expected that 25-40% of the pixels in an agricultural data set will be mixture pixels.)

The data are then processed through a second stage where a pixel is determined to be a pure pixel if the chi-square value for the winning pure case is less than some threshold τ_1. If it is not pure according to this test, then the chi square value for the two class mixture case is compared to a second threshold τ_2. If it is less than τ_2, the pixel is determined to be the mixture indicated; otherwise the pixel is considered to be from a class or classes not included in the signature set.

Currently the thresholds τ_1, τ_2 are chosen empirically -- hence the need for two processing stages. The parameters τ_1 and τ_2 are chosen so as to minimize the error of the proportion estimate over some training area of known proportion.

The signature set described last month was applied to a small 550 pixel section of the data. Subsequent analysis showed that very little of the data were being called out as grass, and as a result the error rate was substantial. It seemed that the initial choice of a grass signature was a poor one (see last month's discussion). Accordingly, a different grass signature was selected, this one being the grass cluster with the most number of pixels. Again the test data were processed through the mixtures classifier. The results were somewhat better, but the total error in the proportion estimation for the test data was still slightly inferior to the error rate achieved using the normal, linear maximum likelihood, classifier. It was further noted that the chi square thresholds chosen, which minimized the total error of the proportion estimate, resulted in 73% of the pixels being counted as "pure" and only 18% of the pixels being assessed as mixtures. Many more mixture pixels had been anticipated. One hypothesis for these results is that the distributions represented by the signatures are both close together and very broad. Thus, perhaps many of the mixture pixels are very near the center of some other distribution. Also, the regular maximum likelihood classification had been done using 15 signatures -- for the mixtures approach we are using only six. It seems that it may be necessary to further pack the signature simplex with other grass signatures so as to increase the grass classification rate.

During the coming month we intend to continue in the analysis of these results and commence further processing.

Signature Extension

One data processing technique to reduce the cost of processing data acquired over either large areas or many similar sites, and to decrease the time required to perform the processing, is by utilizing one set of training data to process all the data collected. Since multispectral scanner data are affected by many natural phenomena which serve to greatly increase the variability in the data[2], the ability to extend the applicability of spectral signatures over a number of data sets or over large variations in the radiation sensed requires the use of special algorithms to adapt the training information to the data or vice versa.

For this contract we intend to develop the capability to extend the applicability of training information for SKYLAB S-192 data.

For this task we have defined a training area and a test area. The training will be carried out over the area in and around Lansing, Michigan. Primary ground classes are dense urban, residential, trees, general vegetation, bare soil, and water. The test area will include Ypsilanti, Michigan and the area immediately to the east. This area has the same types of ground classes as the training area. The sites, which are both included in the S-192 data set we have been working with, are located approximately 80 miles apart. Additionally, the atmospheric state for the two sites is quite different. Examination of S-190B imagery acquired at the same time as the scanner data shows clearly that the training site is cloud free, and the atmosphere appears very clear, while the test site is covered by a considerable haze layer and is surrounded by clouds.

Processing for the signature extension task will include clustering for signatures and subsequent classification for both the training area and the test area. Then the signatures from the training area will be used without any transformation to classify the test site, for use as a bench mark during subsequent analysis. Then the signatures will be modified by various signature extension transformations and used to process the test site data.

The transformations tested will include the MASC algorithm[3], and adaptive classification[4], both developed at ERIM. Further signature extension techniques will be devised and tested as time allows.

Initial processing was begun by clustering an area in the test site. There had been some apprehension that the haze layer would sufficiently affect the reflected radiation that classification could not be adequately carried out. The results of the clustering were good and showed that the area can be well classified.

During the coming month we will continue processing in accordance with the plan outlined above.

References

[1] Horwitz, H. M., et al, Estimating Proportions of Objects From Multispectral Scanner Data, Environmental Research Institute of Michigan, Ann Arbor, ERIM 109600-13-F, May 1975.

[2] Nalepka, R. F. and J. P. Morgenstern, "Signature Extension Techniques Applied to Multispectral Scanner Data", Proceedings of the Eighth International Symposium on Remote Sensing of Environment, Ann Arbor, October 1972.

[3] Henderson, R. G., "Signature Extension Using the MASC Algorithm", Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Ind., June 3-5, 1975.

[4] Crane, R. B., Adaptive Processing With a Decision-Directed Kalman Filter, Environmental Research Institute of Michigan, Ann Arbor, ERIM 190100-31-T, July 1974.

Submitted by: Richard F. Nalepka
Richard F. Nalepka
Principal Investigator

Approved by: Jon D. Erickson
Jon D. Erickson
Head, Information Systems and Analysis Department

Approved by: Paul Rice for
Richard R. Legault
Director, Infrared and Optics Division

dd